
Statistical potentials for fold assessment

FRANCISCO MELO,¹ ROBERTO SÁNCHEZ,² AND ANDREJ SALI

Laboratories of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, The Rockefeller University, New York, New York 10021, USA

(RECEIVED June 12, 2001; FINAL REVISION October 29, 2001; ACCEPTED November 6, 2001)

Abstract

A protein structure model generally needs to be evaluated to assess whether or not it has the correct fold. To improve fold assessment, four types of a residue-level statistical potential were optimized, including distance-dependent, contact, Φ/Ψ dihedral angle, and accessible surface statistical potentials. Approximately 10,000 test models with the correct and incorrect folds were built by automated comparative modeling of protein sequences of known structure. The criterion used to discriminate between the correct and incorrect models was the Z-score of the model energy. The performance of a Z-score was determined as a function of many variables in the derivation and use of the corresponding statistical potential. The performance was measured by the fractions of the correctly and incorrectly assessed test models. The most discriminating combination of any one of the four tested potentials is the sum of the normalized distance-dependent and accessible surface potentials. The distance-dependent potential that is optimal for assessing models of all sizes uses both C_{α} and C_{β} atoms as interaction centers, distinguishes between all 20 standard residue types, has the distance range of 30 Å, and is derived and used by taking into account the sequence separation of the interacting atom pairs. The terms for the sequentially local interactions are significantly less informative than those for the sequentially nonlocal interactions. The accessible surface potential that is optimal for assessing models of all sizes uses C_{β} atoms as interaction centers and distinguishes between all 20 standard residue types. The performance of the tested statistical potentials is not likely to improve significantly with an increase in the number of known protein structures used in their derivation. The parameters of fold assessment whose optimal values vary significantly with model size include the size of the known protein structures used to derive the potential and the distance range of the accessible surface potential. Fold assessment by statistical potentials is most difficult for the very small models. This difficulty presents a challenge to fold assessment in large-scale comparative modeling, which produces many small and incomplete models. The results described in this study provide a basis for an optimal use of statistical potentials in fold assessment.

Keywords: Model evaluation; comparative modeling; fold assignment; fold assessment; statistical potentials; large scale protein structure modeling

Usefulness of a protein structure model depends on its accuracy (Baker and Sali 2001). Thus, it is necessary to estimate accuracy of a three-dimensional (3D) model before it is used. The assessment must generally begin with predict-

ing whether or not the model has at least the correct fold. Such coarse assessment may then be followed by an evaluation of model's detailed features, such as loops and sidechains. The fold has to be assessed almost invariably for all of the models predicted by ab initio methods (Jones 1997; Ortiz et al. 1999; Xia et al. 2000; Bonneau and Baker 2001; Pillardy et al. 2001) and threading of sequences into structures (Bowie et al. 1991; Godzik et al. 1992; Jones et al. 1992; Sippl and Weitckus 1992; Torda 1997). In addition, fold assessment is frequently also needed for models calculated by comparative modeling (Browne et al. 1969; Blundell et al. 1987; Martí-Renom et al. 2000). Comparative or homology modeling builds a model for a protein sequence (target) on the basis of its alignment to known

Reprint requests to: Andrej Sali, 1230 York Avenue, New York, NY 10021, USA; e-mail: sali@rockefeller.edu; fax: (212) 327-7540.

¹Present address: P. Universidad Católica de Chile, Facultad de Ciencias Biológicas, Depto. Genética Molecular y Microbiología, Alameda 340, Santiago, Chile.

²Present address: Structural Biology Program, Department of Physiology and Biophysics, and Institute for Computational Biomedicine, Mount Sinai School of Medicine, Box 1677, 1425 Madison Avenue, New York, New York 10029, USA.

Article and publication are at www.proteinscience.org/cgi/doi/10.1110/ps.22802.

related protein structures (templates). It consists of fold assignment, target-template alignment, model building, and model assessment. Comparative modeling based on <30% sequence identity between the target and the template is common; more than half of all known protein sequences that are detectably related to known protein structures currently share <30% sequence identity with the closest template structure (Pieper et al. 2002). In this low range of sequence similarity, comparative models frequently have an incorrect fold because they are built on an incorrect template structure or a substantially incorrect alignment with the correct template structure. Objective fold assessment is especially important in large-scale, automated comparative modeling of whole genomes in which no user intervention is possible (Martí-Renom et al. 2000). In automated modeling, accurate fold assessment increases the number of the validated models because it identifies the models with the correct fold that are based on statistically insignificant sequence similarity to known protein structures. Fold assessment also increases the average accuracy of the validated models because it weeds out the models with an incorrect fold that are based on statistically significant sequence similarity to unrelated structures or to incorrectly aligned related structures.

A large variety of criteria have been used by the methods for assessing protein structure models. These criteria include deviation from standard bond lengths, bond angles, and dihedral angles (Vriend 1990; Engh and Huber 1991; Morris et al. 1992; Laskowski et al. 1993, 1998) residue or atom packing density (Gregoret and Cohen 1991), molecular mechanics energy functions (Novotny et al. 1984, 1988; Petrey and Honig 2000), distribution of residues between the solvent accessible and buried positions (Bryant and Amzel 1987; Huang et al. 1995), atomic and residue solvation energy (Eisenberg and McLachlan 1986; Baumann et al. 1989; Chiche et al. 1990; Still et al. 1990; Vila et al. 1991; Holm and Sander 1992; Koehl and Delarue 1994; Schaefer et al. 1998; Vorobjev et al. 1998; Cramer and Truhlar 1999; Dominy and Brooks, III 1999; Lazaridis and Karplus 1999; Rapp and Friesner 1999; Gatchell et al. 2000; Kollman et al. 2000; Lee et al. 2000; Petrey and Honig 2000; Wang and Kollman 2000; Zhang et al. 2001), spatial distribution of charged groups (Bryant and Lawrence 1991), distribution of atom–atom distances (Colovos and Yeates 1993), main-chain hydrogen bonding (Laskowski et al. 1993), residue environments (Lüthy et al. 1992; Topham et al. 1994), sequence similarity to related known structures (Sánchez and Sali 1998; Jones 1999), similarity between the secondary structure assignment from the model and secondary structure prediction from the sequence (Jones 1999), atomic volume deviation (Pontius et al. 1996), residue–residue contact area difference (Abagyan and Totrov 1997), occluded surface of residues (Pattabiraman et al. 1995), and a large variety of knowledge-based potentials of mean force or sta-

tistical potentials (Hendlich et al. 1990; Casari and Sippl 1992; Colovos and Yeates 1993; Bauer and Beyer 1994; Kocher et al. 1994; Rooman and Wodak 1995; Jernigan and Bahar 1996; Jones and Thornton 1996; Park and Levitt 1996; Moult 1997; Park et al. 1997; Vajda et al. 1997; Furuichi and Koehl 1998; Melo and Feytmans 1998; Rooman and Gilis 1998; Betancourt and Thirumalai 1999; Rognuckarin and Subramaniam 1999; Lazaridis and Karplus 2000; Tobi and Elber 2000; Tobi et al. 2000; Vendruscolo et al. 2000a). The statistical potentials are generally the single most informative criterion for distinguishing between the models with the correct and incorrect folds, although model assessment may be augmented by simultaneous use of several model features (Jones 1999a, Sánchez and Sali 1999). Statistical potentials are derived from known protein structures and quantify the observed preference of the different residue or atom types to be exposed to the solvent, or to interact with each other in a pairwise or higher order fashion. In addition to assessing of experimentally determined and theoretically predicted protein structures, the statistical potentials have been used in a variety of other applications, including the ab initio protein structure prediction (Sun 1993; Bowie and Eisenberg 1994; O'Donoghue and Nilges 1997; Chiu and Goldstein 2000; Tobi and Elber 2000), fold recognition or threading (Jones et al. 1992; Maiorov and Crippen 1992; Sippl and Weitckus 1992; Bryant and Lawrence 1993; Ouzounis et al. 1993; Huang et al. 1995; DeBolt and Skolnick 1996; Miyazawa and Jernigan 1996, 2000; Reva et al. 1997; Jones 1999b; Kolinski et al. 1999; Panchenko et al. 2000; Skolnick et al. 2000), detection of native-like protein conformations (Hendlich et al. 1990; Casari and Sippl 1992; Bauer and Beyer 1994; Samudrala and Moult 1998; Simons et al. 1999; Gatchell et al. 2000; Vendruscolo et al. 2000b), and prediction of protein stability (Gilis and Rooman 1996, 1997).

In this study, we analyze the effect of the many variables that define the derivation and the use of statistical potentials for discriminating between comparative models with correct and incorrect folds. The tested statistical potentials include distance-dependent, contact, accessible surface and main-chain dihedral angle potentials. The test models with the correct and incorrect folds for proteins of known structure were constructed to be representative of the models obtained from genome-wide comparative modeling calculations, which produce many small, incomplete, and inaccurate models. The criterion used to discriminate between the correct and incorrect models was the Z-score of the model energy. The performance of a given potential was quantified by two tests, the fraction of the correctly assessed test models and the receiver operating characteristic curve.

We begin the Results and Discussion section by characterizing the models used to test various statistical potentials. We continue by describing the performance of the individual and combined potential types. We conclude by sum-

marizing the main lessons learned from the study. In the Materials and Methods section, we describe the derivation of the test models and define all of the tested statistical potentials, as well as the criteria used to assess their performance.

Results and Discussion

Test models

A large set of models is necessary to assess the performance of the statistical potentials in discriminating between the good and bad comparative models. Large sets of good and bad comparative models for proteins of known structure were generated by MODPIPE (Materials and Methods). A good model has the correct fold and is based on a substantially correct alignment; it has >30% of its C_{α} atoms modeled with an error of less than 3.5 Å. On the other hand, a bad model has an incorrect fold or is built on the correct fold using a poor alignment; it has <15% of its C_{α} atoms modeled correctly.

Distributions of several features of the good and bad models are compared in Figure 1. By construction, good models are based on a match with an alignment score higher than 22 nats (Altschul 1998), whereas the bad models are based on matches from 15 to 20 nats. As a result, most of the bad models are based on <30% sequence identity to the template structure, whereas most of the good models are based on <40% sequence identity (Fig. 1A,B). Most of the good and bad models are shorter than 200 residues (Fig. 1C,D). Because a local sequence alignment program was used to generate the alignments needed for modeling, many good and bad models cover only a fraction of the modeled chain. Good models, in general, cover a larger fraction of the modeled chain relative to the bad models (Fig. 1E,F). Most of the good models contain whole domains, whereas most of the bad models correspond to a fraction of a domain only (Fig. 1G,H). Most of the good models have a high percentage of their C_{α} atoms modeled correctly (Fig. 1I,J).

When MODPIPE is applied to whole genomes, distributions of model length and percentage sequence identity are similar to those in Figures 1A–D (Sánchez and Sali 1998). Thus, the current test set of models is probably a good benchmark for the performance of statistical potentials in the assessment of models from genome-wide comparative modeling.

The jack-knife procedure to derive and test the potentials was not used. In other words, the potentials were assessed by using test models, some of which had actual structures of their sequences and of their homologs in the set of known structures from which the potentials were derived. There are four reasons for this apparent lack of statistical rigor.

First, 210 different statistical potentials were calculated and ~10,000 models were assessed with each potential.

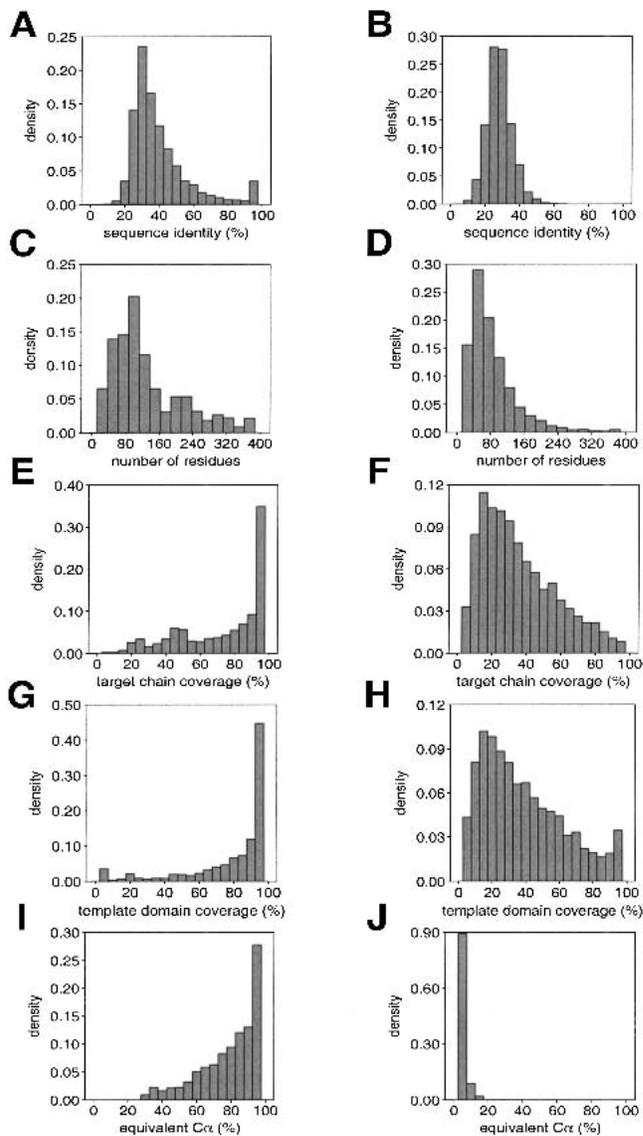


Fig. 1. Properties of the good (*left*) and bad models (*right*). (A,B) Percentage sequence identity between the target and the template. (C,D) Model length. (E,F) Target chain coverage (the fraction of the target chain residues that were modeled). (G,H) Template domain coverage (the fraction of the template domain residues that were aligned to the target chain). The domain coverage was calculated using the domain definitions in the CATH database (Orengo et al. 1997). (I,J) Structural overlap between the target model and the actual target structure expressed as percentage of the equivalent C_{α} atoms (Materials and Methods).

Thus, it was impractical in terms of the CPU time to recalculate each potential for each model assessment. The price for assessing many potentials under many conditions was paid by omitting the jack-knife aspect of the evaluation.

Second, the bias in the evaluation is expected to be small or insignificant, because models with significant errors, not the actual structures, were used for testing.

Third, the main aim was to find an optimal potential for model assessment, not to determine the absolute perfor-

mance of the tested potentials. This aim requires only that the relative performances of the potentials be assessed reliably. The relative performance is expected to be a more robust feature of the learning and test sets of models than the absolute performance.

And finally, it is not certain, even conceptually, that rigorous testing of a method should not rely on structures similar to those from which the potentials were derived. In practice, the statistical potentials are to be used to assess comparative models that by construction are similar to known protein structures. All of the known protein structures are legitimate sources for the statistical potential used in practical model assessment, including those known structures that happen to be related to the assessed model. Thus, it might be that an evaluation of a fold assessment method for comparative modeling that does not eliminate the homologs between the learning and test sets is more accurate even in the absolute sense than an evaluation that does eliminate the overlap. This argument does not apply to evaluating model assessment for the *ab initio* protein structure prediction.

Distance-dependent potentials

Distance-dependent potentials are the most used statistical potentials for fold recognition (Jones et al. 1992; Sippl and Weitckus 1992), protein structure assessment (Sippl 1993; Melo and Feytmans 1998; Jones 1999b), and *ab initio* protein structure prediction (Jones 1997; Xia et al. 2000). In this study, the distance-dependent potentials were derived as described earlier (Sippl 1993; Melo and Feytmans 1998) (Materials and Methods). The critical parameters that define a distance-dependent potential include the range, the resolution (bin size), and the types of the interaction centers. Thus, we explored the effect of varying these parameters on the ability of the distance-dependent potential to discriminate between the good and bad models.

Range

To test the effect of the distance cutoff on the ability of a distance-dependent potential to discriminate between the good and bad models, we derived several potentials by varying only their range from 7 to 50 Å (Fig. 2). The performance of the potential increases with the distance range, depending on the size of the evaluated model. For models larger than 100 residues, the discrimination plateaus at 30 Å. This result is in agreement with prior observations that optimal fold recognition was achieved by distance-dependent potentials cut at 30 Å (Sippl and Jaritz 1994; Furuichi and Koehl 1998). For smaller models, however, this cutoff is ~20 Å. It is easier to discriminate between the good and bad models for large models than for short models. Thus, the size of a model should be taken into account when designing a model assessment method.

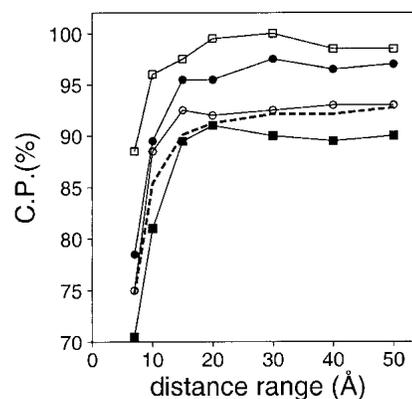


Fig. 2. Performance of the distance-dependent potential as a function of its range. The percentage of the correctly predicted cases for the optimal Z-score cutoff (Materials and Methods). The performance is shown separately for the four sets with 100 good and 100 bad test models each (100/100 sets) (Materials and Methods): The very small models (■), the small models (○), the medium size models (●), and the large models (□). The performance on the 400/400 test model set is indicated by the broken line. The potentials were calculated as specified in Table 1, except for the varying distance range.

Resolution

The resolution of an atomic distance-dependent potential has to be sufficiently high to allow for accurate representation of the different atom type pairs (Melo and Feytmans 1997). In contrast, potentials at the residue level that use only C_{α} atoms, C_{β} atoms or sidechain centers to describe an interaction should require lower resolution for optimal performance. To explore how the resolution influences the performance of a potential, several statistical potentials were derived with bin sizes ranging from 0.1 to 10 Å (Fig. 3). There is a slight improvement in the discriminating power of a potential at a higher resolution for the medium size and large models. In contrast, the small models exhibit a larger

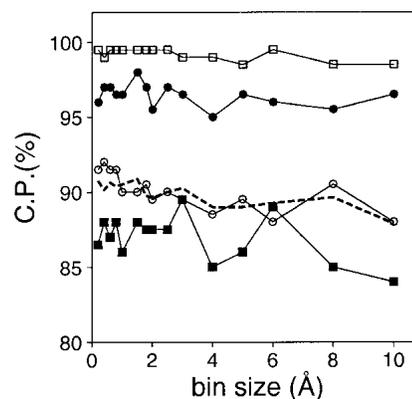


Fig. 3. Performance of the distance-dependent potential as a function of its resolution (bin size). The potentials were calculated as specified in Table 1, except for the varying bin size. See the legend to Fig. 2 for information about the different test model sets represented by the different symbols.

improvement in the performance upon increase in resolution. These results suggest that a resolution higher than 2 Å should be used with residue-level potentials. Most of the distance-dependent potentials described in the literature are within this range of resolution (Sippl 1990; Jones et al. 1992; Lemer et al. 1995; DeBolt and Skolnick 1996; Park and Levitt 1996; Park et al. 1997).

Interaction centers

Traditionally, residue-based statistical potentials are calculated using only C_α atoms, C_β atoms, both C_α and C_β atoms, or centers of the selected mainchain and sidechain atoms (Lemer et al. 1995). In this work, we use only the mainchain and C_β atoms because only a low-resolution representation of protein structure was expected to be sufficient for the main aim, a coarse fold assessment. In addition, the potentials using only the mainchain and C_β atoms have the advantage that the energy Z-score can be calculated rapidly for a large set of sequences with a given structure, because there is no need to rebuild the sidechain atoms. This, in turn, made it possible to explore many potentials under many conditions.

To assess the performance of the distance-dependent potentials with different types of interaction centers, 25 po-

tentials for several combinations of the mainchain (N, C_α , C, O) and C_β atoms were calculated (Fig. 4). The best single atom interaction centers are the C_β atoms. A probable reason is that the C_β atom contains more information about the orientation of the sidechain than any mainchain atom. In contrast, the carbonyl carbon and oxygen atoms have the lowest performances, suggesting that these atoms carry the least amount of information about the sidechain interactions. Using more than one atom per residue slightly improves the performance of the potential. The statistical potentials involving only (C_α , C_β) (Sippl 1993) and (N, O, C_β) (Jones et al. 1992) perform slightly better than the potentials for all the mainchain atoms. The interaction centers for the potentials in Figure 4 are the individual atoms. These potentials performed better than the potentials involving the centers of the selected atoms (data not shown). Thus, the use of the mainchain centers is an unnecessary approximation that leads to a loss of information and to reduced discrimination between the good and bad models by a potential. Apparently, this is not the case when the sidechain centers are used in comparison with the C_α and C_β potentials (Kocher et al. 1994). In this work, however, statistical potentials using sidechain centers of any kind were not assessed because a large-scale calculation of the energy Z-scores with such potentials is impractical.

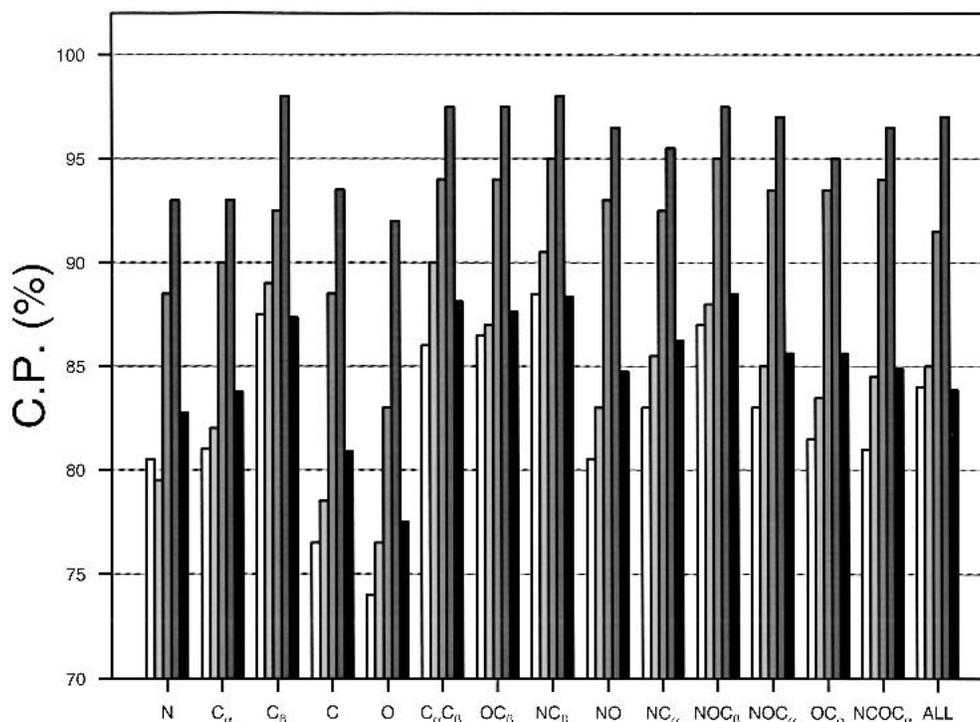


Fig. 4. Performance of the distance-dependent potential as a function of its interaction centers. The atom types whose coordinates were used as the interaction centers are listed on the x-axes. The potentials were calculated as specified in Table 1 except for the varying interaction centers and the potential range of 15 Å. The results for the four 100/100 test sets with models of increasing size are indicated by bars of increasing darkness; the results for the 400/400 set of test models are indicated by the black bars.

Sequence separation

Residues that are close in sequence are restrained to be close in space due to the covalent connectivity between the adjacent residues. Therefore, sequentially local interactions should not be mixed with the sequentially nonlocal interactions when calculating a potential. This distinction should also help to benefit from the specific patterns of interaction in the regular local structures such as α -helices, β -strands, and β -turns. Generally, a potential is extracted and used with the same specification of sequence separation k (Materials and Methods) (Hendlich et al. 1990; Jones et al. 1992; Sippl 1993; Kocher et al. 1994; Park and Levitt 1996; Park et al. 1997; Reva et al. 1997; Rojnuckarin and Subramaniam 1999). However, it is possible to extract a potential only for the nonlocal non-bonded interactions (Melo and Feytmans 1998) and then use it for all of the nonbonded interactions. This approach turned out to be optimal for the ab initio modeling of loops in protein structures (Fiser et al. 2000). Thus, we derived and tested local ($2 < k \leq 8$) and nonlocal ($k \geq 9$) nonbonded potentials separately (Fig. 5).

The best discrimination between the good and bad models is achieved when the local and nonlocal interactions are assessed by potentials derived from the local and nonlocal interactions, respectively (Fig. 5A). The potentials derived from and used to assess only the nonlocal interactions generally perform well, except for the very small models (Fig. 5B). This failure was expected because the number of nonlocal interactions in the very small models is low. Except for the very small models, the nonlocal interactions are more informative than local interactions in all of the tested distance ranges (Fig. 5B,C). An assessment of both local and nonlocal interactions by a potential that was derived only from the nonlocal interactions (Fig. 5D) is better than an assessment of the local (Fig. 5C) or nonlocal (Fig. 5B) interactions on their own.

Known structures for calculating potentials

Two main aspects of the known protein structures used to derive distance-dependent statistical potentials were explored, i.e., their number and their size. The number of

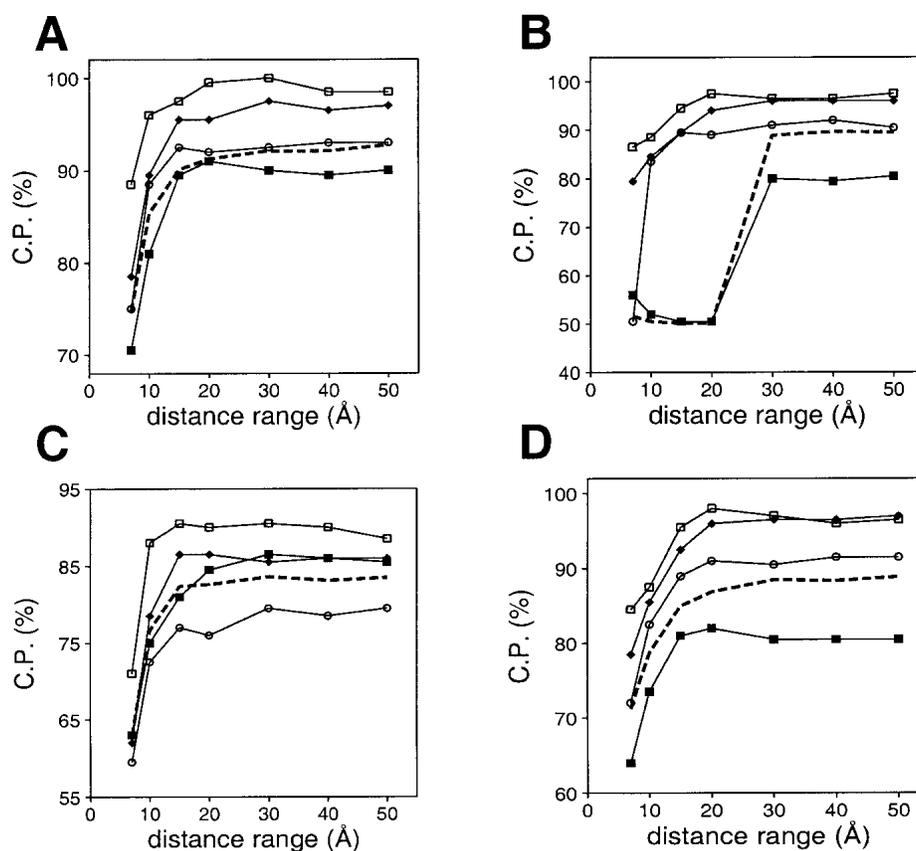


Fig. 5. Performance of the distance-dependent potential as a function of its range and sequence separation. (A) Potentials were derived from and used for assessing both the local ($2 < k \leq 8$) and nonlocal ($k \geq 9$) interactions. (B) Potentials were derived from and used for assessing only the nonlocal interactions. (C) Potentials were derived from and used for assessing only the local interactions. (D) Potentials were derived from the nonlocal interactions, but used to assess both the local and nonlocal interactions, irrespective of their k . See the legend to Fig. 2 for additional information about the potentials and the different test model sets represented by the different symbols.

nonredundant structures in the PDB is limited and some types of interactions may not be sampled densely. Thus, it is important to assess how the performance of a statistical potential depends on the number of protein structures that were used to derive it. To address this question, distance-dependent potentials were derived from 10 sets of known structures containing from 50 to 500 structures (Fig. 6). Discrimination between the good and bad models generally increases with the number of structures used to derive the potential, which is in agreement with previous observations (Sippl and Jaritz 1994; Furuichi and Koehl 1998). However, the improvement of the performance of the potentials upon a 10-fold increase in the number of known structures is small, especially for the medium size and large models, which is consistent with the lack of impact of the database size on the contact potentials (Miyazawa and Jernigan 1996).

The second aspect considered in this study was dependence of model assessment on the size of the known structures used to extract the potential. This exploration was motivated by the previous observations, (1) that statistical potentials depend on the size of the known structures used in their derivation (Sippl 1993; Thomas and Dill 1996; Furuichi and Koehl 1998), although there is some discussion about the generality of this dependence (Bahar and Jernigan 1997), (2) that model size is informative in model evaluation (Sánchez and Sali 1998; Jones 1999b), and (3) that the energy Z-score of a native structure depends strongly on its size (Sippl 1993). Performance of the distance-dependent statistical potential was explored here as a function of the size of the known structures used to derive the potential, its distance range, and size of the assessed models (Fig. 7). For the small and medium size models, discrimination between the good and bad models does not depend significantly on

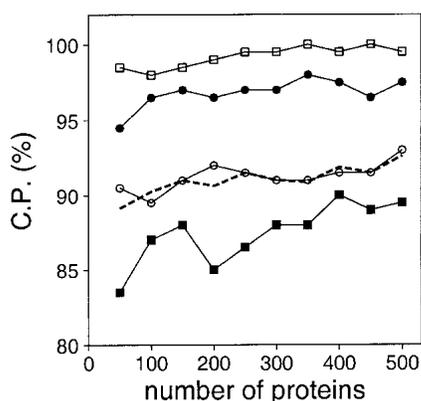


Fig. 6. Performance of the distance-dependent potential as a function of the number of known protein structures used to extract the potential. The potentials were calculated from the 10 sets containing from 50 to 500 known structures (Materials and Methods), as specified in Table 1, except for the potential range of 15 Å. See the legend to Fig. 2 for the different test model sets represented by the different symbols.

the size of the known structures used to extract the potential (Fig. 7B,C). For the very small and large models, it is better to use potentials derived from similarly sized known structures (Fig. 7A,D). For the very small models, this trend becomes weaker when the distance range of the potential increases beyond 15 Å (Fig. 7A). For the large models, the worst discrimination is observed for a potential extracted from the small structures only, especially when the distance range is larger than 15 Å (Fig. 7D), because only a few distant residue–residue contacts occur in small models, thus resulting in highly unfavorable energy values for the distant residue pairs. As such pairs occur frequently in large structures, large models are assessed as unfavorable solely because of their size, not because of their errors. For discriminating between the good and bad models of all sizes, the best performance was achieved by a potential that was derived from known structures of all sizes (Fig. 7E) at all tested distance ranges.

Residue types

In addition to the standard 20 residue types, two residue type definitions were used to calculate a distance-dependent potential, the Wang and Wang residue type group definition that clusters the 20 standard residue types into five groups (Wang and Wang 1999) and the HP model that clusters the 20 standard residue types into two groups according to their hydrophobicity (Huang et al. 1995). For the 400/400 test model set, the fraction of the correctly predicted cases for the distance-dependent potentials using 20 residue types, the HP model, and the Wang and Wang residue type groups were 92.1%, 88.5%, and 84.8%, respectively. Thus, a decrease in the number of residue type groups reduces the discrimination between the good and bad models by the potential.

Subsets of test models

We also compared the predictive power of the distance-dependent statistical potentials by assessing the different subsets of the good and bad models (Materials and Methods). The predictive power of the potentials does not depend strongly on the number of heteroatoms (Materials and Methods), the number of chains, or the percentage sequence identity between the modeled sequences and the templates used to build the assessed models (data not shown). Of the model features tested here, only the size of a model has a major impact on the predictive power of the potentials (see above). Discrimination is more difficult for the small models than it is for the large models. It is tempting to speculate that a model tends to be small when it covers only a fraction of the modeled chain and that the reason for a relatively poor assessment of a small model is its incompleteness; evaluation of a model of an incomplete protein domain is clearly difficult because the environment of the modeled protein fragment is missing. However, the average coverage of the

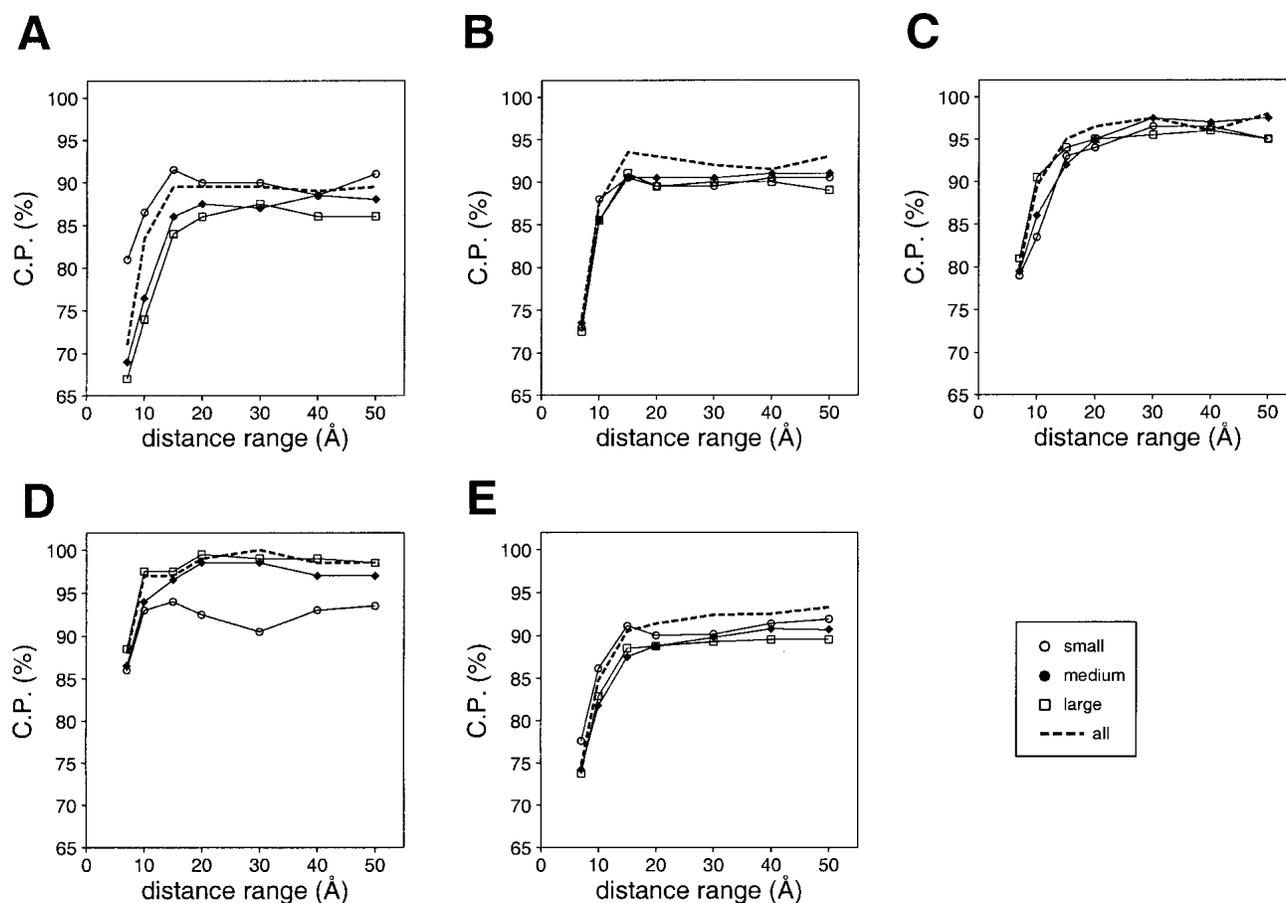


Fig. 7. Performance of the distance-dependent potential as a function of its range and the size of the known structures used to calculate the potential. Four sets of known protein structures were used to extract the potentials: small (<100 residues; ○), medium (100–200 residues; ●), large (>200 residues; ■), and all (the sma-med-large set; broken line) (Materials and Methods). Model assessment by these potentials was evaluated separately for the four 100/100 very small (A), small (B), medium size (C), and large model test sets (D), as well as for the combined 400/400 test set (E) (Materials and Methods). The potentials were calculated as specified in Table 1.

target chain by the correctly assessed models is approximately independent of the model size (data not shown). Thus, the main reason for poor performance in the case of small models is the relatively small number of interactions that are available for discriminating between the good and bad small models. This interpretation is consistent with the previously observed dependence of the Z-score on the size of a model (Sippl 1993). The major challenge in model evaluation remains the assessment of small models.

Contact potentials

The contact potentials are the simplest description of pairwise interactions. They are similar to the distance-dependent potentials, except that they have only two values, the energy of interaction below a contact distance cutoff and zero above it (Miyazawa and Jernigan 1985; DeBolt and Skolnick 1996; Park and Levitt 1996; Park et al. 1997).

Range

Several contact potentials with different distance ranges were calculated and tested (Fig. 8). Similarly to the distance-dependent potentials, discrimination is easier for the large models than it is for the small models. In contrast to the distance-dependent potentials, there is a clear optimal distance range for discriminating between the good and bad models. The optimal distance range of the contact potential increases with the size of the assessed model. Optimal distance values are ~9 Å for the very short models, 11 Å for the small models, 13 Å for the medium size models, and 15 Å for the large models. Thus, contact potentials are less convenient than the distance-dependent potentials for discriminating between the good and bad models, because different contact potentials need to be used to have an optimal discrimination over all model sizes. Performance of the optimal contact potential is comparable with that observed for the distance-dependent potentials only in the case of the

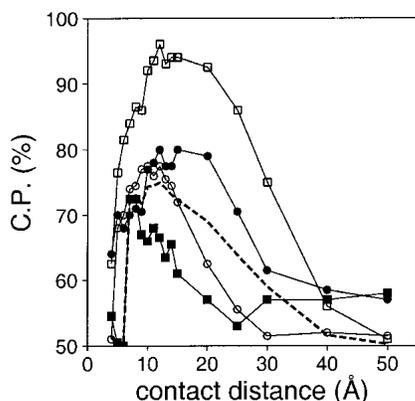


Fig. 8. Performance of the contact potential as a function of its contact distance. The interaction centers were the C_{β} atoms. All the contacts with $k \geq 2$ were considered. The reference state used to calculate the potentials was other residues (Materials and Methods). The potentials were extracted from the sma-med-lar set of known protein structures. See the legend to Fig. 2 for the different test model sets represented by the different symbols.

large models. For the smaller models (<200 residues), the contact potentials are inferior to the distance-dependent potentials. As a result, no further calculations with the contact potentials were performed in this study.

Dihedral angle potentials

Statistical backbone dihedral angle potentials were described previously (Kocher et al. 1994; Gilis and Rooman 1997, 2001). In the current study, the Φ/Ψ dihedral angle potentials for each of the 20 standard residue types (Materials and Methods) did not separate well between the good and bad models, in comparison with the distance-dependent statistical potentials. For example, the fraction of the correctly predicted cases is 67%, 73%, 74%, 84%, and 70% for the four 100/100 very small, small, medium size, and large model sets, and the 400/400 test set, respectively. One of the reasons may be that the relative content of the residues with the native backbone dihedral angles in the bad models is significantly larger than that of the native nonbonded contacts. In other words, the difference between the bad and good models in terms of the backbone dihedral angles may be significantly smaller than the difference in terms of the residue-residue contacts. Because of the poor performance, no further calculations were carried out with the dihedral angle potentials.

Accessible surface potentials

Accessible surface potentials depend on an approximate measure of the residue burial or exposure to the solvent. Accessible surface potentials complement the distance-dependent potentials (Jones et al. 1992; Sippl 1993; Kocher

et al. 1994; O'Donoghue and Nilges 1997; Melo and Feytmans 1998; Jones 1999b). Thus, the accessible surface potentials were assessed here on their own and in combination with other statistical potentials. In this study, the accessible surface potentials were derived as described earlier (Sippl 1993; Melo and Feytmans 1998) (Materials and Methods).

Distance and burial ranges

Two different ranges exist for the accessible surface potentials used in this study, the distance range and the burial range. The distance range is the radius of the sphere surrounding a residue. The burial range is the maximal number of residues observed in the sphere around a residue and corresponds to the minimal solvent exposure. The optimal values for these two ranges were expected to be dependent on each other.

Several accessible surface potentials with varying distance range and a constant burial range were calculated and assessed (Fig. 9). For models larger than 50 residues, good discrimination is achieved over all of the tested distance ranges. The correct prediction rate is approximately constant for potentials with distance ranges from 7 to 15 Å. On the other hand, for the very small models, the rate of the correctly assessed models steadily decreases with the distance range of the potential. Thus, if a single accessible surface potential with good performance over the whole range of protein sizes is required, the distance range should be small (e.g., <9 Å). Whereas a larger distance range results in a small improvement for models larger than 50 residues, it strongly worsens the performance for the very small models.

Accessible surface statistical potentials with a variety of burial ranges were also calculated and evaluated (Fig. 10).

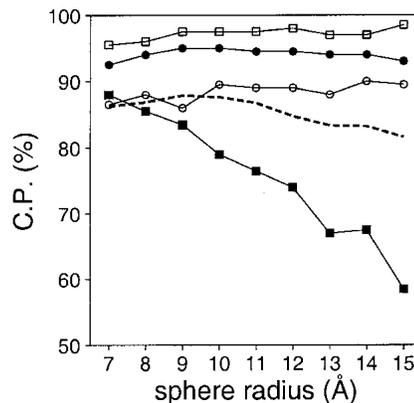


Fig. 9. Performance of the accessible surface potential as a function of its distance range (sphere radius). The potentials were calculated as specified in Table 2, except for the burial range of 200 atoms and the varying sphere radius. See the legend to Fig. 2 for the different test model sets represented by the different symbols.

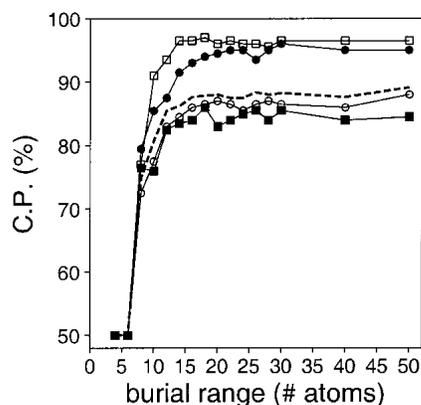


Fig. 10. Performance of the accessible surface potential as a function of its burial range. The potentials were calculated as specified in Table 2, except for the varying burial range. See the legend to Fig. 2 for the different test model sets represented by the different symbols.

The performance of the accessible surface potentials as a function of the burial range is described by an asymptotic curve. These curves reach their maxima at the burial range that depends on the size of the assessed models. For the very small, small, medium size, and large models, the smallest burial range required to reach the maximum discrimination is ~12, 15, 18, and 18 atoms, respectively. Thus, contrary to the distance range, there is a single optimal burial range for assessing models spanning all sizes (i.e., 18 atoms).

Resolution

Several accessible surface potentials with fixed distance and burial ranges but variable burial resolution were calculated (Fig. 11). The performance of the accessible surface potential depends strongly on its resolution, in contrast to that for the distance-dependent potentials (Fig. 3). This dependency varies with the size of the assessed models. For all model sizes, the potential performs best at the high end of the resolution spectrum. The performance falls sharply above a threshold that depends on the model size. The threshold is 5, 8, 13, and 15 atoms for the very small, small, medium size, and large models, respectively.

Interaction centers

The dependence of the performance of the accessible surface potentials on the different interaction centers (Fig. 12) is similar to that of the distance-dependent potentials (Fig. 4). The highest discrimination is obtained for the potential that uses the C_{β} atoms to define the residue accessibility. As pointed out above, it appears that the C_{β} atoms retain more information about the direction of the sidechain and thus better describe residue accessibility than any other single atom type.

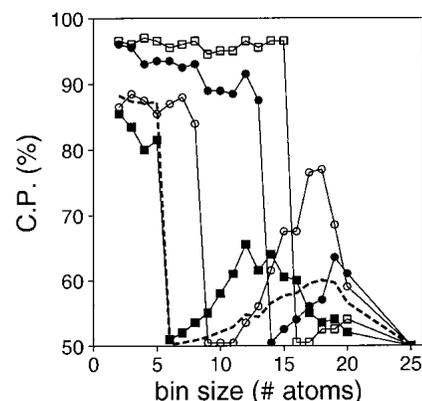


Fig. 11. Performance of the accessible surface potential as a function of its resolution (bin size). The potentials were calculated as specified in Table 2, except for the burial range of 30 atoms and the varying bin size. See the legend to Fig. 2 for the different test model sets represented by the different symbols.

Sizes of assessed model and known structures for calculating potentials

Model assessment by accessible surface potentials was evaluated as a function of the size of the assessed model and the size of the known protein structures used to calculate the potential (Fig. 13). The discriminative ability of a potential increases with the size of the assessed model, similarly to the distance-dependent statistical potentials. The performance is worst for the very small models, presumably for the same reasons as in the case of the distance-dependent potentials. The size of the known structures used to derive the potential influences its discriminant power slightly. Assessment of a model is best by a potential that was derived from known structures of the same size, especially for the very small models.

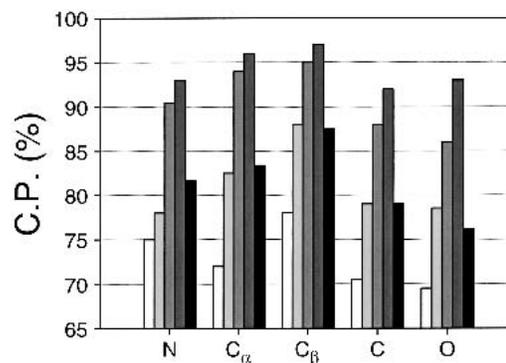


Fig. 12. Performance of the accessible surface potential as a function of its interaction centers. The potentials were calculated as specified in Table 2, except for the distance range of 10 Å and the varying interaction centers. See the legend to Fig. 4 for the different test model sets represented by the different bar shades.

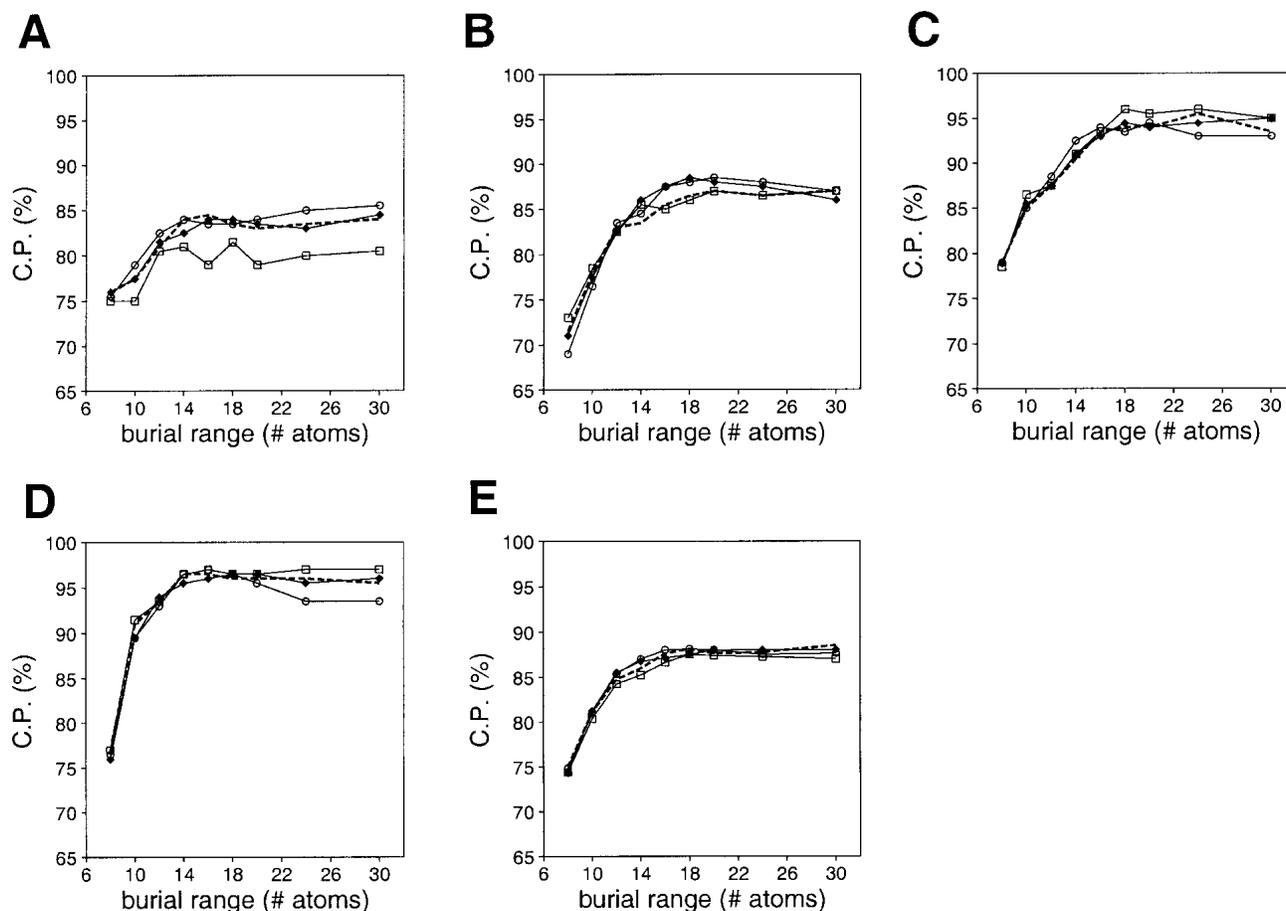


Fig. 13. Performance of the accessible surface potential as a function of its burial range and the size of the known structures used to calculate the potential. Four sets of known protein structures were used to extract the potentials: small (<100 residues; ○), medium (100–200 residues; ●), large (>200 residues; □), and all (the sma-med-large set; broken line) (Materials and Methods). Model assessment by these potentials was evaluated separately for the four 100/100 very small (A), small (B), medium size (C), and large model test sets (D), as well as for the combined 400/400 test set (E) (Materials and Methods). The potentials were calculated as specified in Table 2.

Residue types

The ability of an accessible surface potential to assess models was tested as a function of the three residue-type classifications described above. For the 400/400 test model set, the fraction of the correctly predicted cases for the accessible surface potentials using the 20 standard residue types, the Wang and Wang residue type groups (Wang and Wang 1999), and the HP model (Huang et al. 1995) were 87.5%, 83.8%, and 84.1%, respectively. As for the distance-dependent potential, there is a loss of information when the 20 standard residue types are classified into a smaller number of groups, which can be avoided by using the 20 standard residue types.

Combined potential

A combination of potentials that evaluate different aspects of a model (e.g., residue solvent accessibility and residue–

residue contacts) performs better than the single potentials on their own (Sippl 1993; Kocher et al. 1994). Two specific potentials were combined (Materials and Methods) and tested (Fig. 14), the optimal distance-dependent potential (Table 1) and the optimal accessible surface potential (Table 2). On the 3375/6270 test set of models, the distance-dependent potential performed better than the accessible surface potential, but worse than the combined potential. The specificity and sensitivity of any of the three potentials for the very small models is poor. The difficulty of assessing the very small models is a major hurdle for an improvement of the overall performance. For the medium size and large models, the potentials have low rates of false positives and false negatives (Fig. 14). For example, at the maximum rate of the correct prediction, the fractions of the false positives and false negatives for assessment by the combined potential are 8.9% and 8.5%, respectively. If the structure space reference were used for the calculation of the Z-score in-

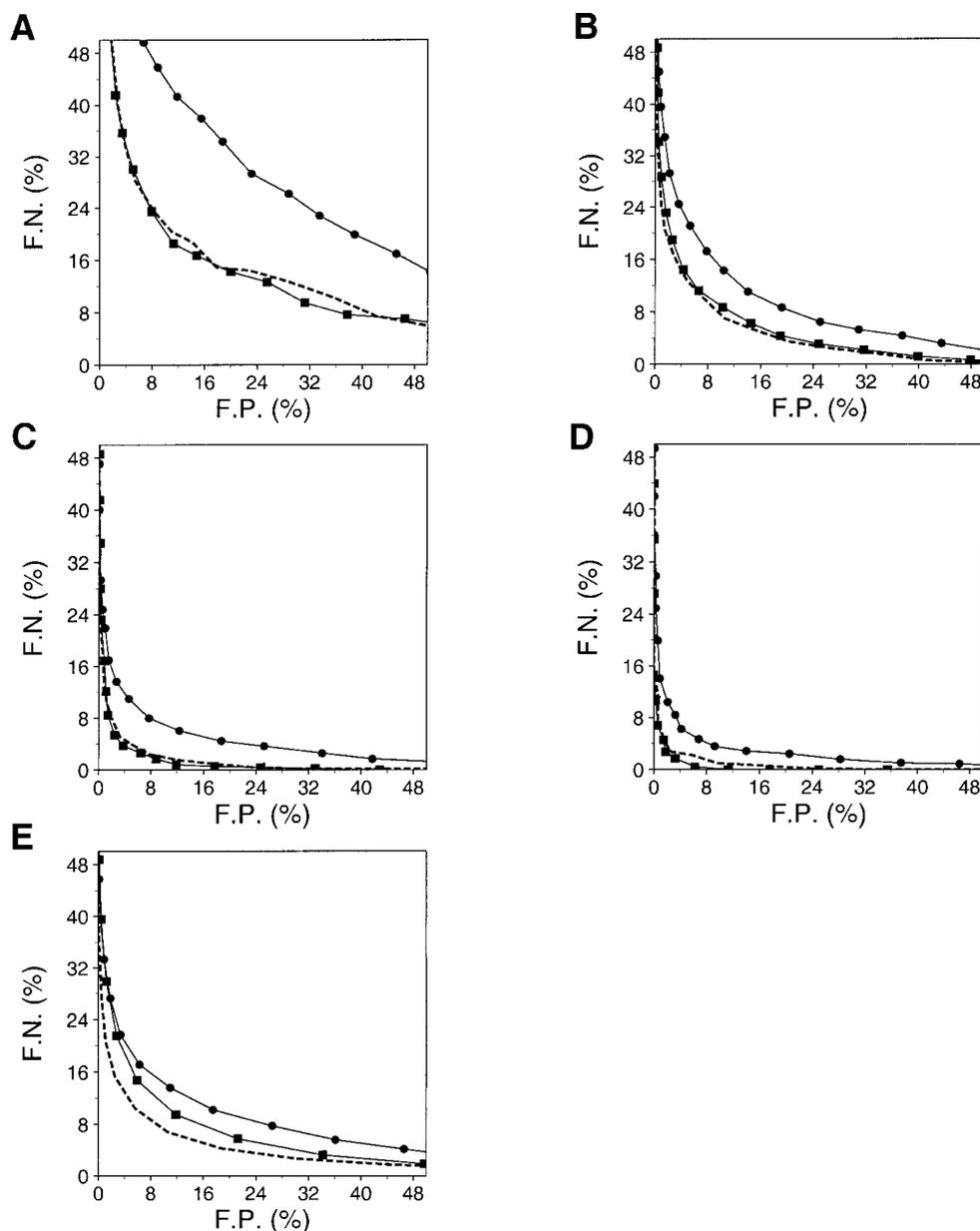


Fig. 14. Performance of the optimal distance-dependent, accessible surface, and combined statistical potentials. The performance is described by the ROC curves, which plot the fraction of false negatives (F.N.) as a function of the fraction of false positives (F.P.) (Materials and Methods). The lower the curve, the better the discrimination between the good and bad models. The ROC curves for the accessible surface potential (●), the distance dependent potential (■), and the combined potential (broken line) are plotted. (A) The 443/1922 test set of the very small models, (B) the 1103/2600 test set of the small models, (C) the 1126/1412 test set of the medium size models, and (D) the 703/336 test set of the large models. (E) The performance of the potentials is also evaluated by the 3375/6270 set of all good and bad models.

stead of the sequence space reference, the performance would be better by approximately two percentage points (Fig. 15). Many different combinations of potentials, including varying weighing of the distance-dependent, contact, Φ/Ψ dihedral angle, and accessible surface potentials, were tested, but none of them performed better than the sum

of the normalized distance-dependent and accessible surface potentials (data not shown).

Conclusions

Four types of a residue-level statistical potential were optimized for fold assessment in large-scale genome-wide com-

Table 1. *The optimized distance-dependent statistical potential*

Parameter	Value
interaction centers	C_α and C_β
residue type classification	20 standard residue types
range	30 Å
resolution	0.50 Å
sequence separation k , derivation	3, 4, 5, 6, 7, 8, ≥ 9
reference state	other atoms
sequence separation k , use	3, 4, 5, 6, 7, 8, ≥ 9

The listed parameters are optimal for all sizes of the tested models. The potential was calculated from the sma-med-lar set of known protein structures, unless specified otherwise. The optimal performance for a given model size is obtained by deriving the potential only from the known structures of similar size.

parative modeling. These potentials included distance-dependent, contact, Φ/Ψ dihedral angle, and accessible surface statistical potentials. The following main conclusions were reached:

1. For an instructive optimization of fold assessment by statistical potentials, the test set of models must be representative of the models to be assessed by the new method. The test set of models used here is believed to be representative of the models from large-scale comparative modeling (Fig. 1).
2. The energy Z-score obtained by randomizing the order of residues in the tested sequence, while keeping the structure constant, is almost as good as that obtained by randomizing the structure, while keeping the sequence constant (Fig. 15).
3. The most discriminating combination of any one of the four tested potentials is that of the distance-dependent and accessible surface potentials (Fig. 14). The distance-dependent potential is more informative than the contact potential, and the Φ/Ψ dihedral angle potential has small discriminative power.
4. The distance-dependent potential that is optimal for assessing models of all sizes uses both C_α and C_β atoms as interaction centers, distinguishes between all 20 standard residue types, has the distance range of 30 Å, resolution of 0.5 Å, and is derived and used by taking into account the sequence separation of the interacting atom pairs (Table 1, Fig. 14). The terms for the sequentially local interactions ($k \leq 8$) are significantly less informative than those for the nonlocal interactions (Fig. 5).
5. The accessible surface potential that is optimal for assessing models of all sizes used C_β atoms as interaction centers, distinguishes between all 20 standard residue types, and has the burial range of 40 atoms (Table 2, Fig. 14). The optimal distance range depends on the model

size, with 9 Å resulting in the best performance averaged over all model sizes.

6. The performance of the tested statistical potentials is not likely to improve significantly with a further increase in the number of known protein structures used in their derivation (Fig. 6).
7. Model size should be taken into account as explicitly as possible when assessing the fold of a model. The parameters of fold assessment whose optimal values vary significantly with model size include the size of the known protein structures used to derive the potential and the distance range of the accessible surface potential (Figs. 7, 9, 13).
8. Fold assessment by statistical potentials is most difficult for the very small models (Fig. 14). Small models are difficult to assess because of the relatively small number of pairwise interactions by which they are judged, not because of their incompleteness. This difficulty presents an important challenge to fold assessment in large-scale comparative modeling, which produces many small models.
9. Attributes of a model other than an energy Z-score may have to be used to improve fold assessments. Such attributes may include model size, fraction of a domain modeled, the significance score of the modeling alignment, and the energy Z-score of the closest template structure.

The results described in this study provide a basis for an optimal use of statistical potentials in fold assessment. They also indicate future directions for the development of more sensitive fold assessment for large-scale comparative modeling (e.g., Conclusions 7 and 9).

Table 2. *The optimized accessible surface statistical potential*

Parameter	Value
interaction centers	C_β
residue type classification	20 standard residue types
distance range	9 Å
burial range	40 atoms
resolution	2 atoms
sequence separation k , derivation	$k \geq 1$
reference state	other atoms
sequence separation k , use	$k \geq 1$

The listed parameters are optimal for all sizes of the tested models, except for the quoted distance range, which corresponds to the optimal average performance over the 400/400 test set of models of all sizes. The potential was calculated from the sma-med-lar set of known protein structures, unless specified otherwise. The optimal performance for a given model size is obtained by deriving the potential only from the known structures of similar size, and by varying the distance range according to Fig. 9.

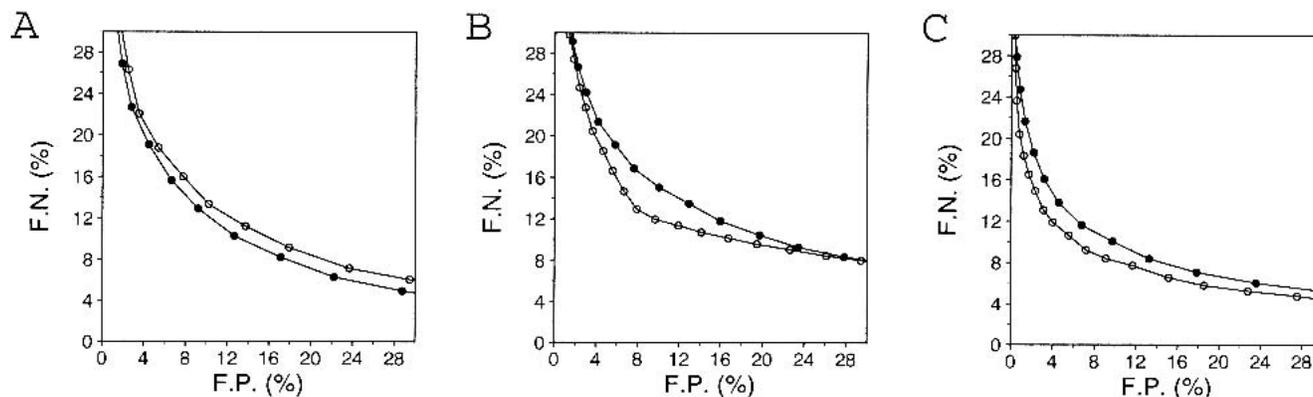


Fig. 15. Performance of the sequence space (●) and structure space (○) references for the calculation of the energy Z-scores. The predictive power is assessed for the 3375/6270 test model set. The statistical potentials and the polyprotein implemented in the program PROSII were used (Sippl 1993). (A) Distance dependent potential. (B) Accessible surface potential. (C) The combined potential.

Materials and Methods

Test models

To evaluate the usefulness of the statistical potentials for fold assessment, large sets of good and bad protein structure models were needed. The two sets of models were calculated by large-scale comparative modeling (Sánchez and Sali 1998) of the protein chains representative of the Protein Data Bank (PDB) of known protein structures (Berman et al. 2000). The models were classified as good or bad depending on their structural similarity to the actual structure of the target protein. All of the models are available at <http://guitar.rockefeller.edu>.

Models with the correct fold (good models)

The good models were built on the basis of the correct templates and mostly correct alignments between the target sequences and the template structures. The models were obtained by applying MODPIPE to 1085 chains representative of the PDB (Sánchez and Sali 1998). These representative sequences corresponded to the protein chains in PDB that shared <30% sequence identity or were >30 residues different in size. The templates for comparative modeling by MODPIPE were 1637 PDB chains with <80% identity to each other or more than 30 residue difference in length. Each target sequence was aligned separately with each one of the 1637 known structures by use of the program ALIGN that implements local sequence alignment by dynamic programming (Altschul 1998). Only the target-template alignments with a significance score higher than 22 nats (corresponding approximately to the PSI-BLAST E-value of 10^{-4}) were used, resulting in 3993 models. Models with <30% structural overlap with the actual experimentally determined structure were eliminated. Structural overlap was defined as the fraction of the equivalent C_{α} atoms upon least-squares superposition of the two structures with the 3.5 Å cutoff. This procedure also removed models based on correct templates that had a poor alignment and models based on templates that had large domain or rigid body movements with respect to the target structure. The final set contained 3375 good models.

Models with an incorrect fold (bad models)

The bad models were built on the basis of a template with an incorrect fold, a template structure with large rigid body shifts, or

an incorrect alignment with the correct template. The models were obtained as described above, except that only the target-template alignments with the significance score between 15 and 20 nats were used; this procedure resulted in 7669 models for the 1085 representative chains. Models with >15% structure overlap with the actual target structure were eliminated. The final set contained 6270 bad models.

Subsets of models

The test set of models containing all of the 3375 good and 6270 bad models is termed the 3375/6270 test set. To analyze the discrimination between the good and bad models by the statistical potentials, the good and bad models were grouped into several subsets on the basis of a variety of criteria. Four 100/100 test sets were created, each one of which contained randomly selected 100 good and 100 bad models of defined size, very-small models (<50 residues), small models (50–100 residues), medium size models (100–200 residues), and large models (> 200 residues). Another test set, the 400/400 set, was created by combining all four 100/100 sets.

All of the good and bad models were also subdivided into the following test sets: very-small models (443/1922), small models (1103/2600), medium size models (1126/1412), and large models (703/336); models based on templates with <40 atoms in nonstandard residue types but water (hetatoms) (2393/5371) and other models (982/899); models based on templates without other chains in the same PDB file (1055/2891) and other models (2320/3379); models based on sequence identity of <30% (890/3722) and other models (2485/2548); models based on sequence identity of <30% and sequence length of >100 residues (614/1578), and models based on sequence identity of >30% and sequence length of more than 100 residues (1188/142).

Known structures for calculating potentials

Protein structures that were solved by X-ray crystallography at a resolution higher than 2.5 Å, with >50 residues, without duplicated or missing atoms and without chain breaks were extracted from the September 1999 version of PDB. Representative structures were selected such that they shared <30% sequence identity with each other or were >30 residues different in length, resulting in the

sma-med-lar set with 760 chains. To assess the dependence of the statistical potentials on the size of the structures from which they were extracted, three subsets of the sma-med-lar set that contained only chains of a certain size were created, the small subset containing protein chains with <100 residues (229 chains), the medium subset containing chains with 100–200 residues (232), and the large subset containing chains with >200 residues (299). To assess the dependence of the statistical potentials on the number of structures from which they were extracted, 10 subsets of the sma-med-lar set were generated randomly, containing 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500 chains. To evaluate the effect of the interactions of a chain with other subunits in a complex, two subsets of the sma-med-lar set were generated, the mon subset containing structures from the PDB files with single chains (273 chains) and the mul set containing chains from the PDB files with more than one chain (225); PDB files with single chains but with a crystal symmetry operator to generate additional chains in the crystal unit cell were excluded. All the PDB sets are available at <http://guitar.rockefeller.edu>.

Statistical potentials

A total of 210 statistical potentials were calculated by varying the following features that define their functional form as follows: (1) type of an interaction center (e.g., individual atoms, gravity centers of several atoms); (2) residue and atom type classification; (3) type of potential (e.g., accessible surface potential, distance-dependent potential); (4) maximum range; (5) bin size for frequency histograms obtained from known structures; (6) for two-residue potentials, sequence separation of residues used to extract the frequency histograms; (7) reference state used to calculate the accessible surface statistical potentials; (8) sequence separation used to calculate the total energy of a model; and (9) set of known protein structures used to extract the frequency histograms. The distinction between features (6) and (8) needs an explanation. A statistical pair potential extracted only from the sequentially nonlocal interactions in the known protein structures (Melo and Feytmans 1998), supplemented by stereochemical restraints, is optimal for describing both the local and nonlocal nonbonded interactions in the modeling of loops (Fiser et al. 2000). Thus, optimizations of the calculation and use of a statistical potential should be done independently from each other. Next, all of the statistical potentials tested in this study are defined in detail.

Distance-dependent statistical potential

The distance-dependent, nonbonded statistical potentials were calculated as described (Sippl 1993; Melo and Feytmans 1997):

$$E_k^{ij}(l) = RT \ln [1 + M_{ijk}\sigma] - RT \ln \left[1 + M_{ijk}\sigma \frac{f_k^{ij}(l)}{f_k^{xx}(l)} \right]. \quad (1)$$

M_{ijk} is the number of occurrences for the interaction center type pair ij separated by k residues in sequence (*i.e.*, $k = |I - J|$, where I and J are the residue indices of interaction center types i and j , respectively):

$$M_{ijk} = \sum_{l=1}^n f(i, j, k, l). \quad (2)$$

n is the number of classes of distances. σ is the weight given to each observation. $\sigma = 1/50$ was used (Sippl 1990), so that with 50

observations $f_k^{ij}(l)$ and $f_k^{xx}(l)$ have equal weights for the calculation of $E_k^{ij}(l)$. $f_k^{ij}(l)$ is the relative frequency of occurrence for the interaction center type pair ij at sequence separation k in the class of distance l .

$$f_k^{ij}(l) = \frac{f(i, j, k, l)}{M_{ijk}}. \quad (3)$$

$f_k^{xx}(l)$ is the relative frequency of occurrence for all the interaction center type pairs at sequence separation k in the class of distance l :

$$f_k^{xx}(l) = \frac{\sum_{i=1}^r \sum_{j=1}^r f(i, j, k, l)}{\sum_{i=1}^r \sum_{j=1}^r \sum_{k=1}^m f(i, j, k, l)} \quad (4)$$

in which r is the number of different interaction center types and m is the number of classes for the sequence separation. The temperature T was set to 293 K, corresponding to RT of 0.582 kcal/mole, where R is the gas constant.

Contact statistical potential

Contact potentials were calculated similarly to the distance-dependent potentials, except for using a single bin with a size equivalent to the range of the potential and considering as equivalent all the interactions between interaction centers with the sequence separation $k \geq 2$.

Accessible surface statistical potential

The accessible surface potentials were calculated as described (Sippl 1993; Melo and Feytmans 1998). The accessible surface of an interaction center is defined as the number of interaction centers within a sphere around the central interaction center; the radius of the sphere is the distance range of the potential. In the case of a known multimeric structure, the neighbor counts were performed only for the interaction centers belonging to the representative chain, but all of the other chains in the PDB file were included in the calculation. From these distributions, the statistical potential was calculated as follows:

$$E^i(r) = RT \ln [1 + M_i\sigma] - RT \ln \left[1 + M_i\sigma \frac{f^i(r)}{f_{ref}^i(r)} \right]. \quad (5)$$

M_i is the frequency of the interaction center type i in all the burial classes:

$$M_i = \sum_{r=1}^R f(i, r), \quad (6)$$

where R is the number of the burial classes and σ represents the weight given to each observation (usually $\sigma = 1/50$). $f^i(r)$ is the relative frequency of occurrence of the interaction center type i in the burial class r :

$$f^i(r) = \frac{f(i, r)}{M_i}. \quad (7)$$

$f_{ref}^i(r)$ is the reference state. Two different reference states were tested. The first one represents the overall relative frequency of

occurrence of the interaction center type i in the burial class r , assuming that the interaction centers are distributed homogeneously over the distance range of the potential. This reference corresponds to an interaction center having no preference for any of the different accessibility states and it can be expressed as:

$$f_{ref}^i(r) = \frac{\left(\frac{M_i}{R}\right)}{M_i} = \frac{1}{R}. \quad (8)$$

Thus, in this case, $f_{ref}^i(r)$ is a constant that depends on the number of burial classes defined for the potential. The second reference state was equivalent to that used for the distance-dependent potentials. Here, $f_{ref}^i(r)$ is the relative frequency of occurrence for all the interaction centers in the burial class r :

$$f_{ref}^i(r) = \frac{\sum_{i=1}^n f(i, r)}{\sum_{i=1}^n \sum_{r=1}^R f(i, r)}, \quad (9)$$

in which n is the number of interaction center types.

Φ/Ψ dihedral angle statistical potential

The Φ/Ψ dihedral angle potentials were obtained for each of the 20 residue types for each of the five areas of the Ramachandran map (Ramachandran et al. 1963). The Ramachandran maps were extracted from a representative set of 1000 protein structures (217,807 residues) that shared <60% sequence identity to each other or had >30-residue length difference, and were determined by X-ray crystallography at a resolution of 2.3 Å or better (Fiser et al. 2000). The Ramachandran plot spanned by the Φ and Ψ dihedral angles was divided into $5^\circ \times 5^\circ$ bins. The frequency of residues in each bin was obtained separately for each of the 20 standard residue types. The Φ/Ψ dihedral angle statistical potential was defined as:

$$E_{\Phi, \Psi}(i) = RT \ln [1 + M_i \sigma] - RT \ln \left[1 + M_i \sigma \frac{f_{\Phi, \Psi}(i)}{f_{\Phi, \Psi}^x} \right], \quad (10)$$

in which σ represents the weight given to each observation (usually, $\sigma = 1/100$). M_i is the number of observations for residue type i in all of the dihedral angle classes:

$$M_i = \sum_{\Phi=1}^n \sum_{\Psi=1}^n f(i, \Psi, \Phi). \quad (11)$$

n is the number of the dihedral angle bins (*i.e.*, 72). $f_{\Phi, \Psi}(i)$ is the relative frequency of occurrence of residue type i with dihedral angles Φ and Ψ :

$$f_{\Phi, \Psi}(i) = \frac{f(i, \Phi, \Psi)}{\sum_{\Phi=1}^n \sum_{\Psi=1}^n f(i, \Phi, \Psi)}. \quad (12)$$

$f_{\Phi, \Psi}^x$ is the reference distribution that corresponds to the relative frequency of occurrence for all residue types with dihedral angles Φ and Ψ :

$$f_{\Phi, \Psi}^x = \frac{\sum_{i=1}^r f(i, \Phi, \Psi)}{\sum_{i=1}^r \sum_{\Phi=1}^n \sum_{\Psi=1}^n f(i, \Phi, \Psi)}, \quad (13)$$

in which r is the number of residue types (the 20 standard residue types were used).

Energy

For distance-dependent potentials, the energy of a protein structure model is a sum of the individual terms over all interaction center pairs i and j , sequence separations k , and distance classes l :

$$E_m = \sum_{i < j, k, l} E(i, j, k, l). \quad (14)$$

For contact potentials, the energy was calculated similarly. For accessible surface and Φ/Ψ dihedral angle statistical potentials, the energy of the protein is equal to the sum of the terms for all of the residues. In addition, the combined energy was defined as the sum of the normalized energies (e.g., the normalized distance-dependent and accessible surface energies). The normalization was achieved by dividing the energy of a certain type (e.g., distance-dependent energy) by the standard deviation of its individual terms.

Energy Z-score

Before an energy was used to discriminate between the good and bad models, it was transformed into a Z-score of energy (Sippl 1993):

$$Z = \frac{E_m - \mu_r}{\sigma_r}, \quad (15)$$

in which E_m is the energy of the model, μ_r and σ_r are the average and standard deviation of the reference energy distribution, respectively. Two different reference energy distributions were calculated and compared. The first approach involved randomization of the order of amino acid residues in the tested model (sequence space reference). This procedure was repeated 200 times, generating 200 models with the same conformation, but different sequences. By construction, the residue composition of these random sequences was identical to that of the original sequence. The energies of the random-sequence models were then used to obtain the average and standard deviation of the reference energy distribution. The second derivation of the reference energy distribution keeps the original sequence, but changes its conformation (structure space reference). The sequence was threaded thousands of times through a set of nonredundant native protein folds (Sippl 1993). This collection of conformations was then used to obtain the average and standard deviation of the energy for the calculation of the Z-score.

The performances of the Z-scores calculated by the two methods were similar, as measured by the receiver operating characteristic curve (see below; Fig. 15). The two reference frames were compared only for the PROSAII statistical distance-dependent, accessible surface and combined potentials (Sippl 1993). The performances of the two reference frames are similar. For the distance-dependent potentials, the sequence space reference was slightly

better than the structure space reference. For the accessible surface potentials, the opposite was the case. For the combined potential, the structure space reference performed slightly better than the sequence space reference. However, for all subsequent calculations, the sequence space reference was used because it is significantly more efficient in terms of the CPU time than the structure space reference.

The fraction of the correctly predicted cases

The fractions of false positives (FP) and false negatives (FN) are defined as:

$$FP = \frac{B}{B + D}, \quad FN = \frac{C}{A + C}, \quad (16)$$

in which A is the number of true positives (good models predicted as good), B is the number of false positives (good models predicted as bad), C is the number of false negatives (bad models predicted as good), and D is the number of true negatives (bad models predicted as bad). The fraction of the correctly predicted cases or the correct classification rate at the optimal value of the Z-score cutoff was used to assess the performance of a given statistical potential in fold assessment as follows:

$$CP = \frac{A' + D'}{A' + B' + C' + D'}, \quad (17)$$

in which the prime is used to indicate the corresponding values at the Z-score cutoff that results in the maximal correct classification rate.

Receiver operating characteristic curves

Receiver operating characteristic (ROC) curves (Theodoridis and Koutroumbas 1999) were also used to assess the statistical potentials. An ROC plot is obtained by plotting the false negatives fraction (1–sensitivity) against the corresponding false positives fraction (1–specificity) for all cutoffs on the Z-score.

The area under the ROC curve represents the probability of incorrect classification over the whole range of cutoffs. This area is usually taken to be an important index because it provides a single measure of overall accuracy that is not dependent upon a particular feature threshold. The area under the curve ranges from 0.0 to 0.5. If it is 0.5, the scores for the good and bad models do not differ (no discriminating power), whereas a value of 0.0 indicates no overlap between the two sets of models (perfect discrimination).

Acknowledgments

We thank Drs. Andr s Fiser, Marc A. Martı-Renom, and Azat Badretdinov for many discussions about the assessment of comparative protein structure models. F.M. was a Norman and Rosita Winston Biomedical Research Foundation Fellow. R.S. was a Howard Hughes Medical Institute predoctoral fellow. A.S. was a Sinsheimer Scholar and an Alfred P. Sloan Research Fellow, and is an Irma T. Hirsch Trust Career Scientist. The investigations have also been aided by grants from NIH (GM 54762) and NSF (BIR-9601845) to A.S. as well as Fundaci n Andes (Project Number 13600/4) to F.M.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Abagyan, R. and Totrov, M. 1997. Contact area difference (CAD): A robust measure to evaluate accuracy of protein models. *J. Mol. Biol.* **268**: 678–685.
- Altschul, S. 1998. Generalized affine gap costs for protein sequence alignment. *Proteins* **32**: 88–96.
- Bahar, I. and Jernigan, R. 1997. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* **266**: 195–214.
- Baker, D. and Sali, A. 2001. Protein structure modeling and structural genomics. *Science* **294**: 93–96.
- Bauer, A. and Beyer, A. 1994. An improved pair potential to recognize native protein folds. *Proteins* **18**: 254–261.
- Baumann, G., Fr mml, C., and Sander, C. 1989. Polarity as a criterion in protein design. *Protein Eng.* **2**: 329–334.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The protein data bank. *Nucleic Acids Res.* **28**: 235–242.
- Betancourt, M. and Thirumalai, D. 1999. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **8**: 361–369.
- Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E., and Thornton, J.M. 1987. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326**: 347–352.
- Bonneau, R. and Baker, D. 2001. Ab initio protein structure prediction: progress and prospects. *Annu. Rev. Biophys. Biomol. Struct.* **30**: 173–189.
- Bowie, J.U. and Eisenberg, D. 1994. An evolutionary approach to folding small α -helical proteins that uses sequence information and an empirical guiding fitness function. *Proc. Natl. Acad. Sci.* **91**: 4436–4440.
- Bowie, J.U., L thy, R., and Eisenberg, D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**: 164–170.
- Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C., and Hill, R.C. 1969. A possible three-dimensional structure of bovine α -lactalbumin based on that of hen's egg-white lysozyme. *J. Mol. Biol.* **42**: 65–86.
- Bryant, S.H. and Amzel, L.M. 1987. Correctly folded proteins make twice as many hydrophobic contacts. *Int. J. Pept. Prot. Res.* **29**: 46–52.
- Bryant, S.H. and Lawrence, C.E. 1991. The frequency of ion-pair substructures in proteins is quantitatively related to electrostatic potential: A statistical model for nonbonded interactions. *Proteins* **9**: 108–119.
- . 1993. An empirical energy function for threading protein sequence through the folding motif. *Proteins* **16**: 92–112.
- Casari, G. and Sippl, M. 1992. Structure-derived hydrophobic potential. Hydrophobic potential derived from x-ray structures of globular proteins is able to identify native folds. *J. Mol. Biol.* **224**: 725–732.
- Chiche, L., Gregoret, L.M., Cohen, F.E., and Kollman, P.A. 1990. Protein model structure evaluation using the solvation free energy of folding. *Proc. Natl. Acad. Sci.* **87**: 3240–3244.
- Chiu, T. and Goldstein, R. 2000. How to generate improved potentials for protein tertiary structure prediction: A lattice model study. *Proteins* **41**: 157–163.
- Colovos, C. and Yeates, T.O. 1993. Verification of protein structures: Patterns of non-bonded atomic interactions. *Protein Sci.* **2**: 1511–1519.
- Cramer, C.J. and Truhlar, D.G. 1999. Implicit solvation models: Equilibria, structure, spectra and dynamics. *Chemical Rev.* **99**: 2161–2200.
- DeBolt, S.E. and Skolnick, J. 1996. Evaluation of atomic level mean force potentials via inverse folding and inverse refinement of protein structures: Atomic burial position and pairwise non-bonded interactions. *Protein Eng.* **9**: 937–955.
- Dominy, B.N. and Brooks, III, C.L. 1999. Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B* **103**: 3765–3773.
- Eisenberg, D. and McLachlan, A.D. 1986. Solvation energy in protein folding and binding. *Nature* **319**: 199–203.
- Engh, R.A. and Huber, R. 1991. Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* **A47**: 392–400.
- Fiser, A., Do, R., and Sali, A. 2000. Modeling of loops in protein structures. *Protein Sci.* **9**: 1753–1773.
- Furuichi, E. and Koehl, P. 1998. Influence of protein structure databases on the predictive power of statistical pair potentials. *Proteins* **31**: 139–149.
- Gatchell, D.W., Dennis, S., and Vajda, S. 2000. Discrimination of near-native protein structures from misfolded models by empirical free energy functions. *Proteins* **41**: 518–534.
- Gillis, D. and Rooman, M. 1996. Stability changes upon mutation of solvent-

- accessible residues in proteins evaluated by database-derived potentials. *J. Mol. Biol.* **257**: 1112–1126.
- . 1997. Predicting protein stability changes upon mutation using database-derived potentials: Solvent accessibility determines the importance of local versus non-local interactions along the sequence. *J. Mol. Biol.* **272**: 276–290.
- . 2001. Identification and ab initio simulations of early folding units in proteins. *Proteins* **42**: 164–176.
- Godzik, A., Kolinski, A., and Skolnick, J. 1992. Topology fingerprint approach to the inverse protein folding problem. *J. Mol. Biol.* **227**: 227–238.
- Gregoret, L.M. and Cohen, F.E. 1991. Effect of packing density on chain conformation. *J. Mol. Biol.* **219**: 109–122.
- Hendlich, M., Lackner, P., Weitckus, S., Floeckner, H., Froschauer, R., Gottsbacher, K., Casari, G., and Sippl, M.J. 1990. Identification of native protein folds amongst a large number of incorrect models. The calculation of low energy conformations from potentials of mean force. *J. Mol. Biol.* **216**: 167–180.
- Holm, L. and Sander, C. 1992. Evaluation of protein models by atomic solvation preference. *J. Mol. Biol.* **225**: 93–105.
- Huang, E.S., Subbiah, S., and M., L. 1995. Recognizing native folds by the arrangements of hydrophobic and polar residues. *J. Mol. Biol.* **252**: 709–720.
- Jernigan, R.L. and Bahar, I. 1996. Structure-derived potentials and protein simulations. *Curr. Opin. Str. Biol.* **6**: 195–209.
- Jones, D. 1997. Successful ab initio prediction of the tertiary structure of n-lysin using multiple sequences and recognized supersecondary structural motifs. *Proteins Suppl* **1**: 185–191.
- . 1999a. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**: 195–202.
- . 1999b. GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287**: 797–815.
- Jones, D.T. and Thornton, J. 1996. Potential energy functions for threading. *Curr. Opin. Str. Biol.* **6**: 210–216.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. A new approach to protein fold recognition. *Nature* **358**: 86–89.
- Kocher, J., Rooman, M., and Wodak, S. 1994. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.* **235**: 1598–1613.
- Koehl, P. and Delarue, M. 1994. Polar and nonpolar atomic environments in the protein core: Implication for folding and binding. *Proteins* **20**: 264–278.
- Kolinski, A., Rotkiewicz, P., Ilkowski, B., and Skolnick, J. 1999. A method for the improvement of threading-based protein models. *Proteins* **37**: 592–610.
- Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., et al. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**: 889–897.
- Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. 1993. Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 283–291.
- Laskowski, R., MacArthur, M., and Thornton, J. 1998. Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.* **5**: 631–639.
- Lazaridis, T. and Karplus, M. 1999. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J. Mol. Biol.* **288**: 477–487.
- . 2000. Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **10**: 139–145.
- Lee, M., Duan, Y., and Kollman, P. 2000. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins* **39**: 309–316.
- Lemer, C., Rooman, M., and Wodak, S. 1995. Protein structure prediction by threading methods: Evaluation of current techniques. *Proteins* **23**: 337–355.
- Lüthy, R., Bowie, J.U., and Eisenberg, D. 1992. Assessment of protein models with three-dimensional profiles. *Nature* **356**: 83–85.
- Maierov, V.N. and Crippen, G.M. 1992. Contact potential that recognizes the correct folding of globular proteins. *J. Mol. Biol.* **227**: 876–888.
- Martí-Renom, M.A., Stuart, A., Fiser, A., Sánchez, R., Melo, F., and Sali, A. 2000. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomolec. Struct.* **29**: 291–325.
- Melo, F. and Feytmans, E. 1997. Novel knowledge-based mean force potential at atomic level. *J. Mol. Biol.* **267**: 207–222.
- . 1998. Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol.* **277**: 1141–1152.
- Miyazawa, S. and Jernigan, R. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**: 623–644.
- . 2000. Identifying sequence-structure pairs undetected by sequence alignments. *Protein Eng.* **13**: 459–475.
- Miyazawa, S. and Jernigan, R.L. 1985. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* **18**: 534–552.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. 1992. Stereochemical quality of protein structure coordinates. *Proteins* **12**: 345–364.
- Moult, J. 1997. Comparison of database potentials and molecular mechanics force fields. *Curr. Opin. Struct. Biol.* **7**: 194–199.
- Novotny, J., Bruccoleri, R., and Karplus, M. 1984. An analysis of incorrectly folded protein models: Implications for structural predictions. *J. Mol. Biol.* **177**: 787–818.
- Novotny, J., Rashin, A.A., and Bruccoleri, R.E. 1988. Criteria that discriminate between native proteins and incorrectly folded models. *Proteins* **4**: 19–30.
- O'Donoghue, S. and Nilges, M. 1997. Tertiary structure prediction using mean-force potentials and internal energy functions: Successful prediction for coiled-coil geometries. *Fold Des.* **2**: S47–S52.
- Orengo, C., Michie, A., Jones, S., Jones, D., Swindells, M., and Thornton, J. 1997. CATH—a hierarchic classification of protein domain structures. *Structure* **5**: 1093–1108.
- Ortiz, A.R., Kolinski, A., Rotkiewicz, P., Ilkowski, B., and Skolnick, J. 1999. Ab initio folding of proteins using restraints derived from evolutionary information. *Proteins* **37**: 177–185.
- Ouzounis, C., Sander, C., Scharf, M., and Schneider, R. 1993. Prediction of protein structure by evaluation of sequence-structure fitness: Aligning sequences to contact profiles derived from three-dimensional structures. *J. Mol. Biol.* **232**: 805–825.
- Panchenko, A., Marchler-Bauer, A., and Bryant, S. 2000. Combination of threading potentials and sequence profiles improves fold recognition. *J. Mol. Biol.* **296**: 1319–1331.
- Park, B. and Levitt, M. 1996. Energy functions that discriminate x-ray and near-native folds from well-constructed decoys. *J. Mol. Biol.* **258**: 367–392.
- Park, B., Huang, E., and Levitt, M. 1997. Factors affecting the ability of energy functions to discriminate correct from incorrect folds. *J. Mol. Biol.* **266**: 831–846.
- Pattabiraman, N., Ward, K., and Fleming, P. 1995. Occluded molecular surface: analysis of protein packing. *J. Mol. Recognit.* **8**: 334–344.
- Petrey, D. and Honig, B. 2000. Free energy determinants of tertiary structure and the evaluation of protein models. *Prot. Sci.* **9**: 2181–2191.
- Pieper, U., Eswar, N., Stuart, A.C., Ilyin, V.A., and Sali, A. 2002. MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res.* (In press).
- Pillardy, J., Czaplowski, C., Liwo, A., Lee, J., Ripoll, D.R., Kazmierkiewicz, R., Oldziej, S., Wedemeyer, W.J., Gibson, K.D., Arnautova, Y.A., et al. 2001. Recent improvements in prediction of protein structure by global optimization of a potential energy function. *Proc. Natl. Acad. Sci.* **98**: 2329–2333.
- Pontius, J., Richele, J., and Wodak, S. 1996. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* **264**: 121–136.
- Ramachandran, G., Ramakrishnan, C., and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain conformations. *J. Mol. Biol.* **7**: 95–99.
- Rapp, C. and Friesner, R. 1999. Prediction of loop geometries using a generalized Born model of solvation effects. *Proteins* **35**: 173–183.
- Reva, B., Finkelstein, A., Sanner, M., and Olson, A. 1997. Residue-residue mean-force potentials for protein structure recognition. *Protein Eng.* **10**: 865–876.
- Rojnuckarin, A. and Subramaniam, S. 1999. Knowledge-based interaction potentials for proteins. *Proteins* **36**: 54–67.
- Rooman, M. and Wodak, S. 1995. Are database-derived potentials valid for scoring both forward and inverted protein folding? *Protein Eng.* **8**: 849–858.
- Rooman, M. and Gilis, D. 1998. Different derivations of knowledge-based potentials and analysis of their robustness and context-dependent predictive power. *Eur. J. Biochem.* **254**: 135–143.
- Samudrala, R. and Moult, J. 1998. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J. Mol. Biol.* **275**: 895–916.
- Sánchez, R. and Sali, A. 1998. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci.* **95**: 13597–13602.
- . 1999. Comparative protein structure modeling in genomics. *J. Comp. Phys.* **151**: 388–401.
- Schaefer, M., Bartels, C., and Karplus, M. 1998. Solution conformations and

- thermodynamics of structured peptides: Molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **284**: 835–848.
- Simons, K., Ruczinski, I., Kooperberg, C., Fox, B., Byströff, C., and Baker, D. 1999. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* **34**: 82–95.
- Sippl, M. 1995. Knowledge-based potentials for proteins. *Curr. Opin. Struct. Biol.* **5**: 229–235.
- Sippl, M. and Jaritz, M. 1994. Predictive power of mean force pair potentials. *Proteins structure by distance analysis*. (ed. H. Bohr and S. Brunak). pp. 113–134 IOS press, Amsterdam, The Netherlands.
- Sippl, M.J. 1990. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213**: 859–883.
- . 1993. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comp. Aid. Molec. Design* **7**: 473–501.
- Sippl, M.J. and Weitckus, S. 1992. Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. *Proteins* **13**: 258–271.
- Skolnick, J., Kolinski, A., and Ortiz, A. 2000. Derivation of protein-specific pair potentials based on weak sequence fragment similarity. *Proteins* **38**: 3–16.
- Still, W.C., Tempezyk, A., Hawley, R.C., and Hendrickson, T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**: 6127–6129.
- Sun, S. 1993. Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. *Protein Sci.* **2**: 762–785.
- Theodoridis, S. and Koutroumbas, K. 1999. *Pattern recognition*. Academic Press.
- Thomas, P. and Dill, K. 1996. Statistical potentials extracted from protein structures: How accurate are they? *J. Mol. Biol.* **257**: 457–469.
- Tobi, D. and Elber, R. 2000. Distance-dependent, pair potential for protein folding: Results from linear optimization. *Proteins* **41**: 40–46.
- Tobi, D., Shafran, G., Linial, N., and Elber, R. 2000. On the design and analysis of protein folding potentials. *Proteins* **40**: 71–85.
- Topham, C.M., Srinivasan, N., Thorpe, C.J., Overington, J.P., and Kalsheker, N.A. 1994. Comparative modelling of major house dust mite allergen *der p* I: Structure validation using an extended environmental amino acid propensity table. *Protein Eng.* **7**: 869–894.
- Torda, A.E. 1997. Perspectives in protein-fold recognition. *Curr. Opin. Struct. Biol.* **7**: 200–205.
- Vajda, S., Sippl, M., and Novotny, J. 1997. Empirical potentials and functions for protein folding and binding. *Curr. Opin. Struct. Biol.* **7**: 228–228.
- Vendruscolo, M., Mirny, L., Shakhnovich, E., and Domany, E. 2000a. Comparison of two optimization methods to derive energy parameters for protein folding: Perceptron and Z score. *Proteins* **41**: 192–201.
- Vendruscolo, M., Najmanovich, R., and Domany, E. 2000b. Can a pairwise contact potential stabilize native protein folds against decoys obtained by threading? *Proteins* **38**: 134–148.
- Vila, J., Williams, R.L., Vásquez, M., and Scheraga, H.A. 1991. Empirical solvation models can be used to differentiate from near-native conformations of bovine pancreatic trypsin inhibitor. *Proteins* **10**: 199–218.
- Vorobjev, Y.N., Almagro, J.C., and Hermans, J. 1998. Discrimination between native and intentionally misfolded conformations of proteins: Es/Is, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins* **32**: 399–413.
- Vriend, G. 1990. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**: 52–56.
- Wang, J. and Wang, W. 1999. A computational approach to simplifying the protein folding alphabet. *Nat. Struct. Biol.* **6**: 1033–1038.
- Wang, W. and Kollman, P. 2000. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* **303**: 567–582.
- Xia, Y., Huang, E., Levitt, M., and Samudrala, R. 2000. Ab initio construction of protein tertiary structures using a hierarchical approach. *J. Mol. Biol.* **300**: 171–185.
- Zhang, L., Gallicchio, E., Friesner, R., and Levy, R.M. 2001. Solvent models for protein-ligand binding: Comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations. *J. Comp. Chem.* **22**: 591–607.